Journal of Biogeography (J. Biogeogr.) (2010) 37, 1797–1810

ORIGINAL ARTICLE

Predicting invasive alien plant distributions: how geographical bias in occurrence records influences model performance

René Wolmarans, Mark P. Robertson* and Berndt J. van Rensburg

Department of Zoology and Entomology, Centre for Invasion Biology, University of Pretoria, Pretoria 0002, South Africa

ABSTRACT

Aim To investigate the impact of geographical bias on the performance of ecological niche models for invasive plant species.

Location South Africa and Australia.

Methods We selected 10 Australian plants invasive in South Africa and nine South African plants invasive in Australia. Geographical bias was simulated in occurrence records obtained from the native range of a species to represent two scenarios. For the first scenario (A, worst-case) a proportion of records were excluded from a specific region of a species' range and for the second scenario (B, less extreme) only some records were excluded from that specific region of the range. Introduced range predictions were produced with the MAXENT modelling algorithm where models were calibrated with datasets from these biased occurrence records and 19 bioclimatic variables. Models were evaluated with independent test data obtained from the introduced range of the species. Geographical bias was quantified as the proportional difference between the occurrence records from a control and a biased dataset, and environmental bias was expressed as either the difference in marginality or tolerance between these datasets. Model performance [assessed using the conventional and modified AUC (area under the curve of receiver-operating characteristic plots) and the maximum true skill statistic] was compared between models calibrated with occurrence records from a biased dataset and a control dataset.

Results We found considerable variation in the relationship between geographical and environmental bias. Environmental bias, expressed as the difference in marginality, differed significantly across treatments. Model performance did not differ significantly among treatments. Regions predicted as suitable for most of the species were very similar when compared between a biased and control dataset, with only a few exceptions.

Main conclusions The geographical bias simulated in this study was sufficient to result in significant environmental bias across treatments, but despite this we did not find a significant effect on model performance. Differences in the environmental spaces occupied by the species in their native and invaded ranges may explain why we did not find a significant effect on model performance.

Keywords

Australia, ecological niche modelling, environmental bias, geographical bias, invasive alien plants, marginality, MAXENT, model performance, South Africa, tolerance.

*Correspondence: Mark P. Robertson, Department of Zoology and Entomology, Centre for Invasion Biology, University of Pretoria, Pretoria 0002, South Africa.
E-mail: mrobertson@zoology.up.ac.za

Journal of Biogeography

INTRODUCTION

Natural ecosystems, overall biodiversity, human health and natural economies are affected at global, regional and local scales by an ever-increasing number of destructive alien species (Vitousek *et al.*, 1997; Pimentel *et al.*, 2001). Correlative models that quantify the relationship between occurrence records and environmental data have been applied to a number of problems in biology (see Guisan & Zimmermann, 2000; Pearson & Dawson, 2003). These models have not only been particularly useful for predicting species distributions and ecological niches of invasive organisms (Robertson *et al.*, 2001, 2003; Steiner *et al.*, 2008), they have also provided valuable insight into the potential spread of invasive organisms (Peterson, 2003; Rouget *et al.*, 2004; Richardson & Thuiller, 2007) and have predicted the environmental suitability of regions that have not yet been invaded (Mgidi *et al.*, 2007; De Meyer *et al.*, 2008). Predictions such as these can be used to identify areas where management and monitoring efforts should be focused.

Most models have attempted to predict the distribution of invasive species using the relationship between occurrence records from the introduced range (adventive range) of the species and environmental data (Rouget *et al.*, 2004). This relationship assumes that a species' current distribution provides useful information regarding the species' environmental requirements (Pearson *et al.*, 2007). The implicit assumption of this approach is that these species are in equilibrium with their environment. Hence, enough time has passed since their introduction to allow for all the environmentally suitable sites to be occupied. This may not be the case for all species (Wilson *et al.*, 2007); for example sleeper weeds that have a long lag phase between introduction and naturalization (Guisan & Thuiller, 2005). In cases where a species has only recently started to invade a new region, or for which invasion risk needs to be assessed (Mgidi *et al.*, 2007; De Meyer *et al.*, 2008), it may be preferable or necessary to make predictions using native range records (in an attempt to quantify the niche of an organism). Recent studies have indicated that niches can differ between native and introduced ranges of invasive species (Broennimann *et al.*, 2007; Fitzpatrick *et al.*, 2007; Beaumont *et al.*, 2009), prompting the use of occurrence records from both native and introduced ranges in models (Mau-Crimmins *et al.*, 2006; Broennimann & Guisan, 2008; Steiner *et al.*, 2008; Beaumont *et al.*, 2009).

Geographical bias often exists in datasets of collection records due to biased sampling. Biased sampling can be the result of collectors mainly focusing their attention on specific areas, e.g. areas that are easily accessible (near roads or rivers) (Funk & Richardson, 2002) or areas with high species richness and diversity (Dennis & Thomas, 2000; Loiselle *et al.*, 2008). When occurrence records are geographically biased, the underlying environmental gradients in which a species can persist will most likely also not be fully sampled, which could result in environmental bias (Raes & ter Steege, 2007; Hortal *et al.*, 2008). The records available from herbaria or electronic databases may only sample part of a species' geographical range or, in some cases, sampling effort in certain parts of the range may be better than others (Raes & ter Steege, 2007). This is likely to be the case when a species has a native range that spans several countries. In addition, acquisition of all the records can be time-consuming and expensive, and also difficult if sources are in different languages (Mau-Crimmins *et al.*, 2006). If parts of the native range of a species cannot be sampled, or are poorly sampled, then the potential distribution of the species in the introduced range is likely to be underestimated. This would occur if the areas that were not sampled contain portions of the environmental gradients in which the species can persist that were not represented with the records obtained from the sampled areas (Welk, 2004; Guisan & Thuiller, 2005; Hortal *et al.*, 2008).

The aim of this study was to examine the impact of geographical bias on the performance of ecological niche models of invasive plant species. We examine models that were calibrated with native range occurrence records and utilized to predict the potential distribution in the introduced range. Different treatments, consisting of biased datasets, were created with the native range occurrence records to simulate different geographical bias scenarios. Specifically, we wanted to test the influence of the resultant underlying environmental bias on model performance, and in so doing we addressed the following questions. (1) What is the relationship between geographical bias and environmental bias in the native range of a species? (2) Does environmental bias differ across treatments that simulate geographical bias? (3) Is there sufficient environmental bias to result in significant differences in model performance across the treatments?

MATERIALS AND METHODS

Species selection and distribution records

We selected South Africa and Australia as the study regions since both countries have exchanged several plant species that have become naturalized, and in many cases invasive (Randall, 2002; Henderson, 2007). Both countries have good sources of distribution records that are readily available through electronic databases. We compiled a list of 10 Australian plants that are invasive alien species in South Africa and a list of nine South African species that are invasive alien species in Australia (Table 1). For the Australian species invading South Africa, native range occurrence records were obtained from the Australian Virtual Herbarium public access database (AVH; http://www.ersa.edu.au/avh/, accessed 15 February 2007). Additional occurrence records were requested from the Queensland Herbarium and the Western Australia Herbarium (FloraBase database). Introduced range occurrence records were obtained from the Southern African Plant Invaders Atlas (SAPIA; http://www.agis.agric.za/, accessed 15 February 2007; Henderson, 2007).

For the South African species invading Australia, native range occurrence records were obtained from South Africa's

Table 1 Species selected for analysis. The first ten species are Australian species that have invaded South Africa and the following nine (indicated with an asterisk) are South African species that have invaded Australia. The number of native range records (occupied cells) used for calibration and the number of introduced range records (occupied cells) used for evaluation are given.

Botanical name	Family	Common name	Number of native records	Number of introduced records
Acacia cyclops	Fabaceae	Red-eye	100	150
Acacia dealbata	Fabaceae	Silver wattle	200	285
Acacia decurrens	Fabaceae	Green wattle	101	110
Acacia longifolia	Fabaceae	Sydney golden wattle	220	86
Acacia mearnsii	Fabaceae	Black wattle	135	430
Acacia melanoxylon	Fabaceae	Australian blackwood	287	138
Acacia saligna	Fabaceae	Port Jackson willow	268	146
Hakea sericea	Proteaceae	Silky hakea	47	79
Leptospermum laevigatum	Myrtaceae	Australian myrtle	112	36
Paraserianthes lophantha	Fabaceae	Stinkbean	83	45
*Arctotheca calendula**	Asteraceae	Cape weed	97	431
*Ehrharta calycina**	Poaceace	Veld grass	240	115
*Moraea flaccida**	Iridaceae	One leaf Cape tulip	16	41
*Oxalis pes-caprae**	Oxalidaceae	Soursob	39	117
*Polygala myrtifolia**	Polygalaceae	Sweet pea bush	125	41
*Romulea rosea**	Iridaceae	Oniongrass	57	109
*Sparaxis bulbifera**	Iridaceae	Sparaxis	17	58
*Watsonia meriana**	Iridaceae	Bulbil watsonia	44	36
*Zantedeschia aethiopica**	Araceae	Arum lily	60	34

National Herbarium Computerised Information System (PRECIS) database. Introduced range occurrence records were derived from the AVH, Queensland Herbarium and Western Australia herbarium.

In both cases, all native range records were assumed to be representative of the entire native range of the species concerned. As a large proportion of the records were collected using a 15′ grid, all model calibration and evaluation was done at this spatial resolution. The coordinates of the occurrence records were used to assign each record to a grid cell in the map region. Only one record per grid cell was used when several occurrence records were present in a grid cell (Hernandez *et al.*, 2006).

Environmental predictors

We selected the 19 bioclimatic variables available from the WorldClim database (http://www.worldclim.org; Hijmans *et al.*, 2005; see Appendix S1 in Supporting Information). Bioclimatic variables derived from monthly temperature and precipitation data are commonly used in biogeographical modelling (De Meyer *et al.*, 2008; Loiselle *et al.*, 2008). These variables represent annual trends (e.g. annual mean temperature), seasonality (e.g. annual range in precipitation) and extreme or limiting environmental factors (e.g. precipitation of the wettest month) (Hijmans *et al.*, 2005). All environmental predictors were resampled to 15′ grids using ArcGIS 9.2.

Simulating biased sampling

We started by defining the full range of a species as all of the native range distribution records available for that species. To simulate geographical bias we defined two scenarios that could occur. In the first scenario (A) a proportion of the distribution records were excluded from a specific geographical region of the full range of a species (Fig. 1a). This represents an extreme case where no records are available for the species from a specific part of the range. This could occur for species with ranges that include a politically unstable country in which no records have been collected for that species. The second scenario (B) is less extreme than the first because only some records were excluded from that region (Fig. 1b). This simulates a case where sampling effort is poor in a particular part of the range. For each species we generated three datasets for each scenario. For scenario A we excluded 10%, 20% and 40% of the records from the north or west of the range [whichever showed the greatest change in annual precipitation (variable BIO 12)] to produce three datasets A10, A20 and A40. For scenario B, an initial sample of 50% of the records from either the east or south of the range was used and then a further 40% of the remaining records (from the west or north, respectively) was randomly sampled and added to the initial sample (B10). This was repeated by adding 30% and 10%, respectively, to the initial sample of 50% to generate the remaining two datasets (B20 and B40). We also generated datasets by randomly excluding 10%, 20% and 40% from the full range (Fig. 1c) to serve as controls (R10, R20 and R40). Datasets with the same number in the code (e.g. A10, B10 and R10) contained equal numbers of occurrence records.

Ecological niche modelling

Maxent is a relatively new statistical modelling technique that has been applied to model the potential distribution of species and to estimate niche occupation (Phillips *et al.*, 2006; Peterson *et al.*, 2008; Phillips & Dudík, 2008). An advantage

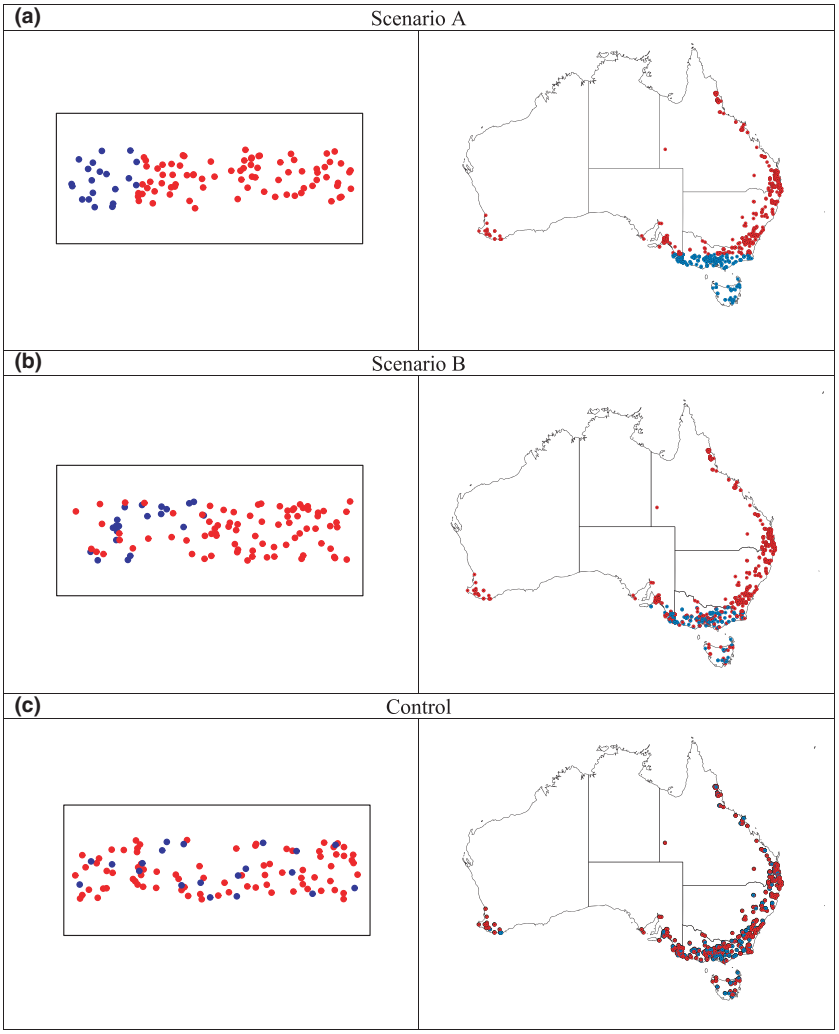

Figure 1 A hypothetical example indicating the selection of occurrence records used to simulate geographical bias in datasets that were used to calibrate models. Blue symbols indicate occurrence records that were excluded and red symbols indicate those that were selected. (a) Illustrating a geographical bias scenario (scenario A) where a proportion of records (e.g. 10%; A10) were excluded from a specific geographical region of a species' range. (b) Illustrating a geographical bias scenario (scenario B) where only some records were excluded (e.g. 10%; B10) from a specific region of a species' range. (c) Indicates random exclusion of occurrence records from the entire region of the species' range for the control datasets (e.g. 10%; R10).

of this approach is that it uses presence records to contrast the distribution of the occurrence sites with the environmental conditions of the entire area of interest or a random background sample of sites, also referred to as pseudo-absences (Pearson *et al.*, 2007). For most invasive species, occurrence data have been recorded without a planned sampling design. Therefore the majority of occurrence records consist of presence-only data obtained from herbarium collections (Soberón & Peterson, 2005; Elith *et al.*, 2006). Even when absence data are available, they are usually unreliable, as the species may have been recorded as absent merely because insufficient time had elapsed to allow for invasion (Wilson *et al.*, 2007; Peterson *et al.*, 2008). Maxent probability responses can be raw, logistic or cumulative, and for this study we selected the logistic probability response following advice given by Phillips & Dudík (2008). This is a continuous variable ranging from 0 to 1, where high values indicate higher suitability for a species in a particular grid cell (Phillips *et al.*, 2006).

All models were calibrated with samples of records taken from the native range (the calibration set) (Fig. 2). Environmental variable values corresponding to each grid cell of the calibration set were extracted for each species and used as input data for Maxent (version 3.1.0). Algorithm parameters were set to a maximum number of 500 iterations, a regularization multiplier of 1, auto features and a convergence threshold of 0.00001.

We wanted to ensure that native range predictions were accurate and acceptable before making projections to the introduced range. Five native range predictions were made; in each case all the native range occurrence records obtained were partitioned into a calibration set (training set) and a testing set (validation set) using a *k*-fold method (Pearson *et al.*, 2006; Phillips *et al.*, 2006). A random selection of 70% of the occurrence records comprised the calibration set and the remaining 30% comprised the testing set. For the testing set, pseudo-absences were generated by randomly selecting grid cells from the map region that did not contain presence records. For each species the number of pseudo-absence records was equal to the number of presence records in the testing set (Jiménez-Valverde & Lobo, 2007; McPherson & Jetz, 2007; Roura-Pascual *et al.*, 2008). Models projected on to the introduced range were evaluated with a test set that was comprised of occurrence records available for each species

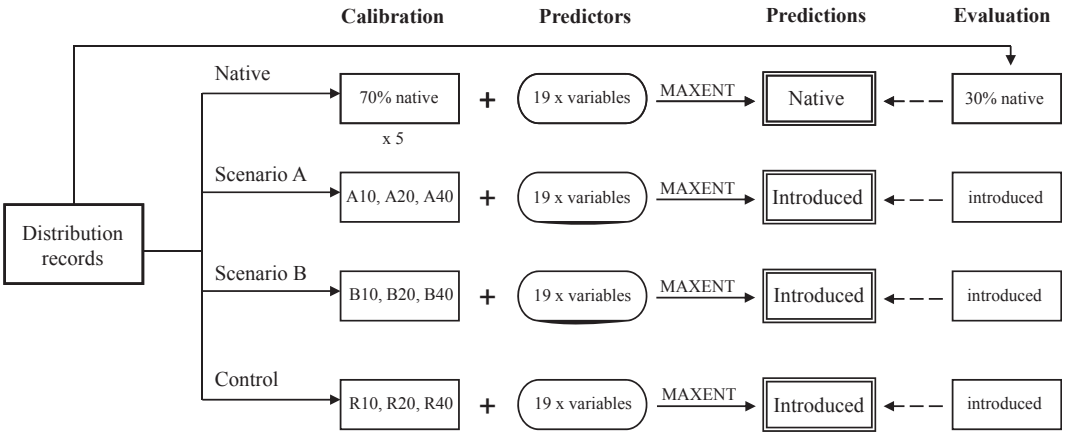

Figure 2 A flowchart of the different datasets used to make native and introduced range predictions for a species. All datasets comprise samples of occurrence records obtained from the native range of a species.

from the introduced range and pseudo-absence records generated in a similar fashion as described above.

Ecological niches

We used ecological-niche factor analysis (ENFA; Hirzel *et al.*, 2002) implemented in the adehabitat package in R (R Development Core Team, 2004) to compare the ecological niche occupied by a species in its native and introduced range (Steiner *et al.*, 2008). ENFA compares the prevailing climatic conditions in a specific range with background data that represent environmental conditions globally (Steiner *et al.*, 2008). We used the same 19 bioclimatic variables derived from the WorldClim database to serve as global environmental background data, to allow for direct comparisons among species (Steiner *et al.*, 2008). ENFA characterizes two components of a species' ecological niche, namely marginality and tolerance. Marginality measures the niche position by comparing the mean of a specific range with the background mean, and tolerance measures the niche width by comparing the variance in a specific range with the background environmental variation. A difference in the marginality values indicate a shift in the mean environmental niche occupied by a species, and a difference in the tolerance values indicates a shift in the niche width (Steiner *et al.*, 2008). To determine whether niche shift occurred between the native and introduced range of a species, we compared marginality values between the native and introduced range with Student's *t*-tests, and contrasted tolerance values with Bartlett's homogeneity of variance tests (Steiner *et al.*, 2008).

In addition, we wanted to assess visually whether each species occupied the same environmental space in the native and introduced range, since models were calibrated with native occurrence records and evaluated with records obtained from the introduced range. We performed a principal components analysis (PCA) on the environmental data extracted from native and introduced datasets for each species (Robertson *et al.*, 2001; Mau-Crimmins *et al.*, 2006) and plotted the native records used for calibration and the introduced records used for evaluation in the resulting environmental space. This allowed us to establish whether the records used for evaluation and for calibration overlapped in environmental space.

Model performance was examined for models that predicted the native range of a species and models that were projected to the introduced range. This allowed us to determine how well the models could predict a species' known distribution in the native range and could therefore verify that models were calibrated with a reliable set of predictor variables.

Model evaluation

Model performance is frequently assessed using the area under the curve (AUC) of receiver-operating characteristic (ROC) plots (Fielding & Bell, 1997; Lobo *et al.*, 2008; Peterson *et al.*, 2008). All AUC values range between 0 and 1, where values of 0.5 indicate that predictions are no better than random, predictions are poor when values range between 0.5 and 0.7 and useful if values fall between 0.7 and 0.9. Predictions that obtain values greater than 0.9, are considered to be good or even excellent (Fielding & Bell, 1997; Broennimann *et al.*, 2007). Recently, the use of AUC statistics for model evaluation has been criticized (Lobo *et al.*, 2008). A modified AUC approach was proposed by Peterson *et al.* (2008) to overcome some of these criticisms. This method plots the true positive fraction as a function of the proportion of the overall area predicted present. This evaluation method eliminates the reliance on commission error, where areas might be classified as unsuitable based on uncertain absences, i.e. pseudo-absences (Peterson *et al.*, 2008). These ROC results are expressed as ratios of the area under the observed curve to the area under the random line. We selected three model evaluation statistics, including two threshold-independent statistics (the AUC and modified AUC) and one threshold-dependent statistic (the true skill statistic, TSS). The AUC statistic was selected as this is the most commonly used threshold-independent statistic and will allow comparison with other studies. We calculated the modified

AUC statistic due to weaknesses in the AUC statistic. Additionally, we used the TSS instead of the popular Kappa statistic, as Kappa is influenced by prevalence (Allouche *et al.*, 2006). TSS values range from −1 to +1, where +1 indicates a perfect fit and values of 0 or less indicate a performance no better than random (Allouche *et al.*, 2006). Since the TSS value was calculated for all possible thresholds ranging from 0 to 1 with set increments of 0.01, only the maximum TSS (mTSS) value for each species prediction was reported. All three model performance measures were calculated using functions written in R (R Development Core Team, 2004).

Analysis

In order to describe the relationship between geographical and environmental bias we had to quantify these biases. We quantified environmental bias as the difference in either the marginality (niche position) or tolerance value (niche width) between a control (e.g. R10) and bias dataset (e.g. A10 or B10). The marginality and tolerance value for a dataset comprising biased records is likely to be smaller than for a dataset comprising randomly selected records, as a biased dataset will tend to sample less environmental variation. We quantified geographical bias as the proportional difference between a control and bias dataset using the *x*- or *y*-coordinates of a particular dataset, depending on the direction from which records were excluded in order to simulate bias. We show how the bias created in the occurrence records relates to the geographical and environmental distribution of the species (see Appendix S2). To assess whether the amount of environmental bias differed on average across treatments we compared the difference in marginality or difference in tolerance values for all the species between a control dataset and the relevant bias dataset (e.g. R10 and A10). Nonparametric comparisons were carried out with Kruskal–Wallis tests, and then the significant differences were identified with multiple comparisons carried out with the npmc function in R (R Development Core Team, 2004). We describe the relationship between environmental bias and model performance by determining the correlation between the difference in marginality or tolerance and difference in model performance between a control (e.g. R10) and bias dataset (e.g. A10) across all treatments. We show the potential distribution predicted for two species in the introduced range with the different biased and corresponding control datasets.

RESULTS

Ecological niches

Comparisons of marginality values between the native and introduced range indicated that there was no significant difference in the position of the niche occupied between these two ranges for any of the species (Table 2). However, a comparison of the tolerance values between these two ranges indicated that the environmental variation between the native and introduced range differed significantly for 12 of the 19 species (Table 2).

An assessment of the PCA analyses revealed that for 9 out of the 19 species (*Acacia cyclops*, *Acacia mearnsii*, *Acacia saligna*, *Arctotheca calendula*, *Moraea flaccida*, *Oxalis pes-caprae*, *Romulea rosea*, *Sparaxis bulbifera* and *Watsonia meriana*), the environmental space occupied by the species in the introduced range expanded in relation to the environmental space occupied in the native range (Table 2, Appendix S2b). In addition, 5 out of the 19 species (*Acacia dealbata*, *Acacia longifolia*, *Acacia*

Table 2 Marginality and tolerance values compared between the native and introduced range of the study species. Marginality values were contrasted with Student's *t*-tests and tolerance values with Bartlett's test for homogeneity. The first ten species are Australian species that have invaded South Africa and the following nine (indicated with an asterisk) are South African species that have invaded Australia.

Species	Native marginality	Introduced marginality	Student's *t*-test	Native tolerance	Introduced tolerance	Bartlett's test
Acacia cyclops	1.29	1.33	0.82	3.94	2.55	**20.37**
Acacia dealbata	1.39	1.22	4.48	2.42	2.31	0.49
Acacia decurrens	1.24	1.26	0.50	2.49	6.23	**79.01**
Acacia longifolia	1.37	1.30	1.14	1.79	2.22	**5.32**
Acacia mearnsii	1.35	1.19	3.88	2.46	1.96	**9.85**
Acacia melanoxylon	1.36	1.19	3.42	1.88	2.09	2.12
Acacia saligna	1.25	1.31	1.37	2.38	2.26	0.54
Hakea sericea	1.38	1.35	0.51	3.50	3.14	0.65
Leptospermum laevigatum	1.46	1.48	0.31	2.09	4.63	**24.67**
Paraserianthes lophantha	1.38	1.42	0.77	2.79	4.77	**14.23**
*Arctotheca calendula**	1.20	1.16	0.69	1.82	1.82	0
*Ehrharta calycina**	1.25	1.24	0.24	1.74	3.24	**49.44**
*Moraea flaccida**	1.34	1.27	3.78	34.64	8.45	**25.54**
*Oxalis pes-caprae**	1.31	1.20	1.66	2.86	2.31	2.41
*Polygala myrtifolia**	1.26	1.37	2.01	2.03	4.65	**30.21**
*Romulea rosea**	1.27	1.26	0.23	2.90	2.50	1.51
*Sparaxis bulbifera**	1.35	1.34	0.56	27.82	6.66	**28.36**
*Watsonia meriana**	1.31	1.38	1.43	8.21	3.48	**26.66**
*Zantedeschia aethiopica**	1.23	1.37	2.07	2.07	5.73	**32.50**

Values indicated in bold were significantly different ($P < 0.05$) between the different ranges.

melanoxylon, *Leptospermum laevigatum* and *Zantedeschia aethiopica*) showed a smaller environmental space occupied in the introduced range relative to the environmental space occupied in the native range (Table 2, Appendix S2b). For all species, native range models obtained AUC values (with conventional ROC analysis) that ranged between 0.94 and 1.00, indicating good/excellent model performance (Appendix S3). Models projected onto the introduced range showed more variation in performance (Appendix S3). Models for nearly all the species achieved AUC values greater than 0.8, indicating that model performance in the introduced range was useful. However, *Acacia decurrens*, *A. dealbata* and *A. mearnsii* obtained values close to and below 0.7 (Appendix S3), indicating useful but poor performance in the introduced range.

The relationship between geographical bias and environmental bias

The environmental space occupied with all the native range records compared with the environmental space occupied when 40% of the records were excluded from the dataset (A40 treatment, biased records) was similar for some species (e.g. *A. decurrens* in Fig. 3b, *Hakea sericea*, *Paraserianthes lophantha*, *Ehrharta calycina* and *Polygala myrtifolia*), but for the majority of species the environmental space occupied with these two datasets was different (Appendix S2). There was considerable variation in the relationship between geographical bias and environmental bias across treatments (Fig. 4). For some species, an increase in geographical bias led to an increase in the amount of environmental bias, but no clear overall trends were evident. The correlation between the proportional geographical difference and marginality difference was only significant for the B10 treatment (Table 3); the rest of the treatments showed no significant correlation (Table 3). None of the treatments showed a significant correlation when environmental bias was expressed as the difference in tolerance.

Differences in environmental bias across treatments

Environmental bias, expressed as the difference in marginality between treatment and control datasets for all species (Appendix S4a), differed significantly across treatments ($H = 26.13$, $P < 0.05$). Nonparametric multiple comparisons (multiple Steel tests) identified that these differences were especially large in the comparisons of A10 and A40 ($P = 0.004$) as well as B10 and B40 ($P = 0.057$). When measured as the difference in tolerance (Appendix S4b), environmental bias showed no significant differences across treatments ($H = 10.18$, $P = 0.069$).

Environmental bias and model performance across treatments

Model performance did not differ significantly among treatments with different levels of bias (Table 4). Some species (e.g. *A. decurrens*, *A. longifolia* and *H. sericea*) had a large difference

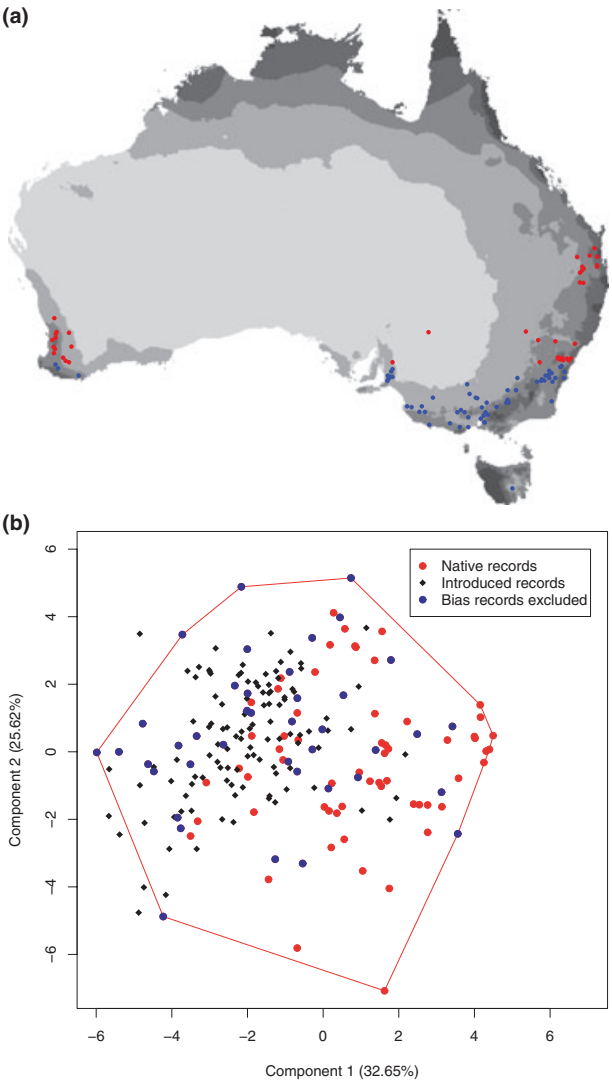

Figure 3 A diagrammatic illustration of the relationship between (a) the geographical and (b) the environmental bias for *Acacia decurrens*. The geographical distribution of the species is shown across the annual precipitation gradient (BIO 12) encountered in the native range (darker colours indicate wetter areas). The extent of the entire native range of the species is indicated with red dots and the geographical distribution is represented with the biased dataset (A40) indicated with blue dots. The environmental distribution of the species is shown by the principal components analysis (PCA) of the environmental conditions associated with the native (red dots) and introduced (black dots) range of the species. The blue dots represent the records excluded when the A40 dataset was created. The percentage of environmental variation explained by a specific component is given in parentheses.

in model performance measured between the control and bias dataset, even though there was only a small amount of environmental bias present in the bias dataset (Fig. 5). The converse was also true. Nearly all treatments showed a positive correlation between the difference in marginality and the difference in AUC, although the relationships were not significant (Table 5). Similarly, all treatments showed a

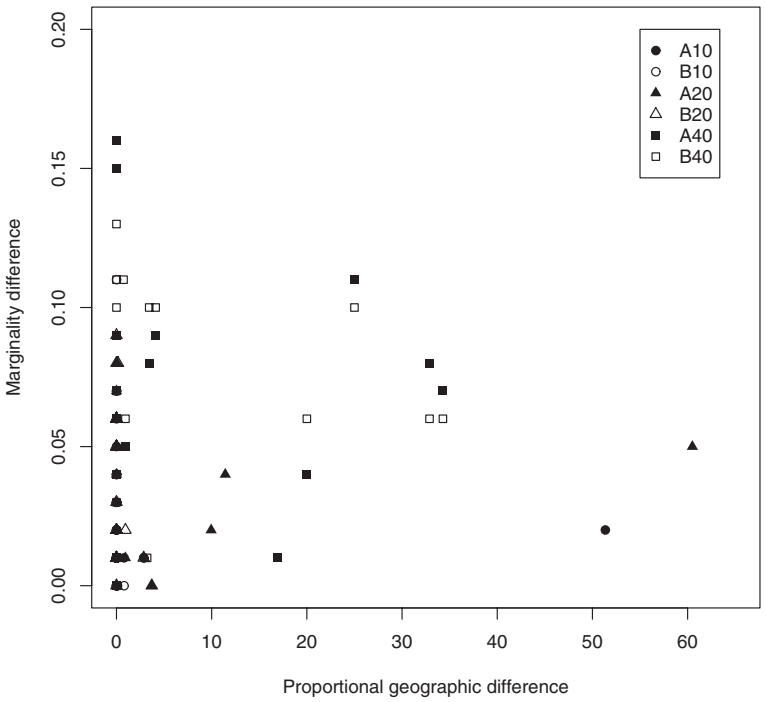

Figure 4 Relationship between geographical and environmental bias assessed across all treatments for all species. Geographical bias is quantified as the proportional difference between records from a control and bias dataset using the *x*- or *y*-coordinates of a particular dataset, depending on the direction from which records were excluded in order simulate bias (measured in decimal degrees). Environmental bias is expressed as the difference in marginality between a control and bias dataset.

Table 3 Relationship between geographical bias (expressed as proportional geographical difference) and environmental bias in the datasets for 19 invasive plant species (expressed as the difference in marginality or tolerance) evaluated across all treatments. Regression (R^2-value) and correlation (*r*-value) analyses were carried out by comparing values from a control (e.g. R10) and bias dataset (A10 or B10).

	Difference in marginality			Difference in tolerance		
Treatment	R^2	$F(1,17)$	*r*	R^2	$F(1,17)$	*r*
R10, A10	0.000	0.004	−0.019	0.026	0.450	−0.101
R10, B10	0.069	1.250	**−0.510**	0.012	0.199	−0.322
R20, A20	0.033	0.588	−0.060	0.095	1.789	−0.464
R20, B20	0.171	3.510	−0.285	0.000	0.005	−0.150
R40, A40	0.063	1.140	0.168	0.000	0.002	−0.154
R40, B40	0.000	0.007	−0.021	0.000	0.013	−0.140

$F(1,17)$ is the proportion of variation explained between the control and bias group, degrees of freedom (n–2), $n = 19$. No significant relationship was found ($P > 0.05$), except for the B10 treatment indicated in bold.

Table 4 Comparisons of the performance of models for 19 invasive plant species compared across all treatments using Kruskal–Wallis tests. Model performance was measured with the conventional area under curve analysis (AUC), maximum true skill statistic (mTSS) and the modified version of the AUC analysis (Aratio).

Model performance	Treatments		
	R10, A10, B10	R20, A20, B20	R40, A40, B40
AUC	0.005	0.545	0.024
mTSS	0.062	0.182	0.022
Aratio	0.012	0.292	0.028

Values of the test statistic obtained when each model performance measure was compared between a control and biased datasets ($P > 0.05$).

positive correlation between the difference in tolerance and the difference in AUC; a significant relationship was evident for the A40 and B40 treatments (Table 5). The region predicted as suitable for *A. saligna* in the introduced range was very similar when compared between a biased (e.g. A10 or B10) and control dataset (e.g. R10) across all treatments (see Appendix S5). This is one of the species that shows very little difference in performance across treatments. Range predictions for *A. dealbata*, however, were rather different when the ranges predicted with control datasets were compared with ranges predicted with the biased datasets (Fig. 6). The differences were especially large when the prediction obtained with the R40 dataset was compared with the ranges obtained with the A40 and B40 dataset. This is one of the species that showed a large difference in performance across treatments.

DISCUSSION

Relationship between geographical bias and environmental bias

We found considerable variation in the relationship between geographical bias and environmental bias when examined across species. This result is consistent with that of Kadmon *et al.* (2004) and suggests that the effect of geographical bias on the underlying environmental bias most likely depends on where the geographical range of the species is located in

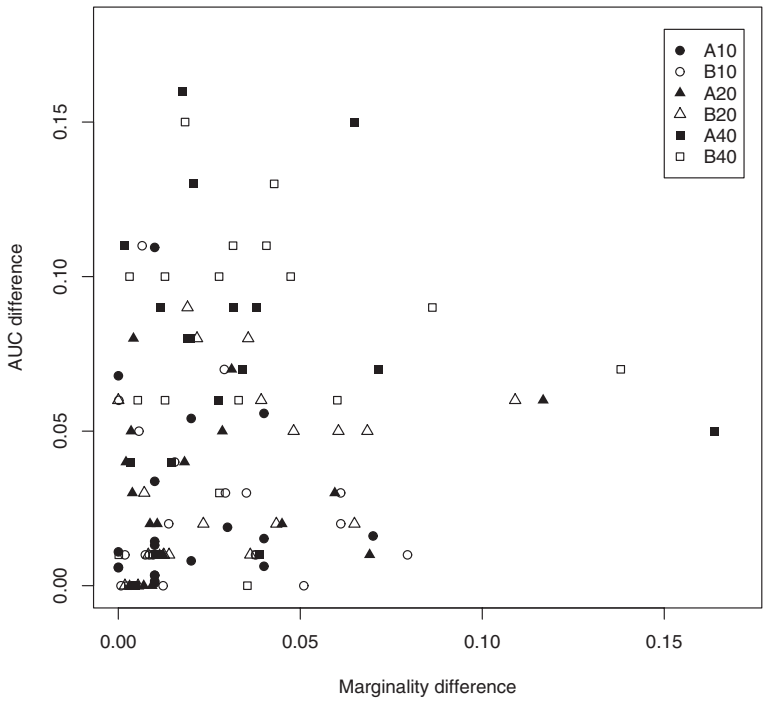

Figure 5 The relationship between environmental bias and model performance assessed across all treatments for all species. Environmental bias is expressed as the difference in marginality. Model performance is expressed as the AUC [area under the curve obtained from the conventional receiver-operating characteristic (ROC) analysis] difference calculated as the absolute difference in AUC values between a specific control (e.g. R10) and bias dataset (e.g. A10).

Table 5 Correlation between environmental bias and model performance assessed for different datasets simulated for 19 invasive plant species. Environmental bias was expressed as the difference in marginality or tolerance. Model performance was assessed by comparing a control (e.g. R10) and bias dataset (A10 or B10).

	Difference in marginality		Difference in tolerance	
Treatment	*r*	*P*-value	*r*	*P*-value
R10, A10	0.250	0.301	0.015	0.952
R10, B10	−0.082	0.738	0.046	0.852
R20, A20	0.131	0.594	0.204	0.403
R20, B20	0.330	0.168	0.190	0.436
R40, A40	0.199	0.414	0.578	**0.009**
R40, B40	0.229	0.346	0.633	**0.004**

Values indicated in bold were significantly different ($P < 0.05$).

relation to the underlying environmental gradients and the steepness of these gradients. For example, a species with a geographical range that covers a steep environmental gradient can be expected to show considerable environmental bias in the presence of a small amount of geographical bias in the dataset. We focused on the annual precipitation gradient in both native ranges when describing the main environmental gradient encountered in a range, as this variable was identified to have the largest amount of variation. In Australia this gradient gradually changes from very dry interior areas to wetter areas all along the coast. Species native to Australia showed geographical distributions all along the coastal areas that varied from the south-western coast to the south-eastern coast excluding Tasmania (*A. cyclops*), the south-western to eastern coast including Tasmania (*A. dealbata*, *A. longifolia*, *A. mearnsii* and *Paraserianthes lophantha*), the south-western coast all along to the east coast excluding Tasmania (*A. decurrens*), the north-eastern to the south-western coast including Tasmania (*A. melanoxylon*), the west to east coast excluding Tasmania (*A. saligna*), the south-western coast to east coast including Tasmania (*Leptospermum laevigatum*), and an east/south eastern coast only distribution (*Hakea sericea*). In South Africa, however, low-rainfall areas are encountered in the north-western side of the country and gradually increase towards the eastern and north eastern parts of the country (O'Brien, 1993). Species native to South Africa showed geographical distributions that stretched from the west coast to the interior part of the country to the south-eastern coast (*Arctotheca calendula*), all along the coast to the northern interior (*Ehrharta calycina*, *Polygala myrtifolia* and *Zantedeschia aethiopica*), along the south-western coast only (*Moraea flaccida* and *Sparaxis bulbifera*), the west/south western coast (*Oxalis pes-caprae*) and from the south to the south-western coast (*Romulea rosea*). We show that the geographical bias that was simulated in the native range occurrence records translated into different amounts of environmental bias across species (Appendix S2). For some species the environmental space occupied by all the native range records and the environmental space occupied by the records that were excluded from the dataset (biased records) was similar. In other words, a small amount of environmental bias existed between the two datasets. For the majority of the species, however, different parts of the environmental space were occupied by records from these two datasets, indicating larger amounts of environmental bias. We did not find any significant relationship between geographical and

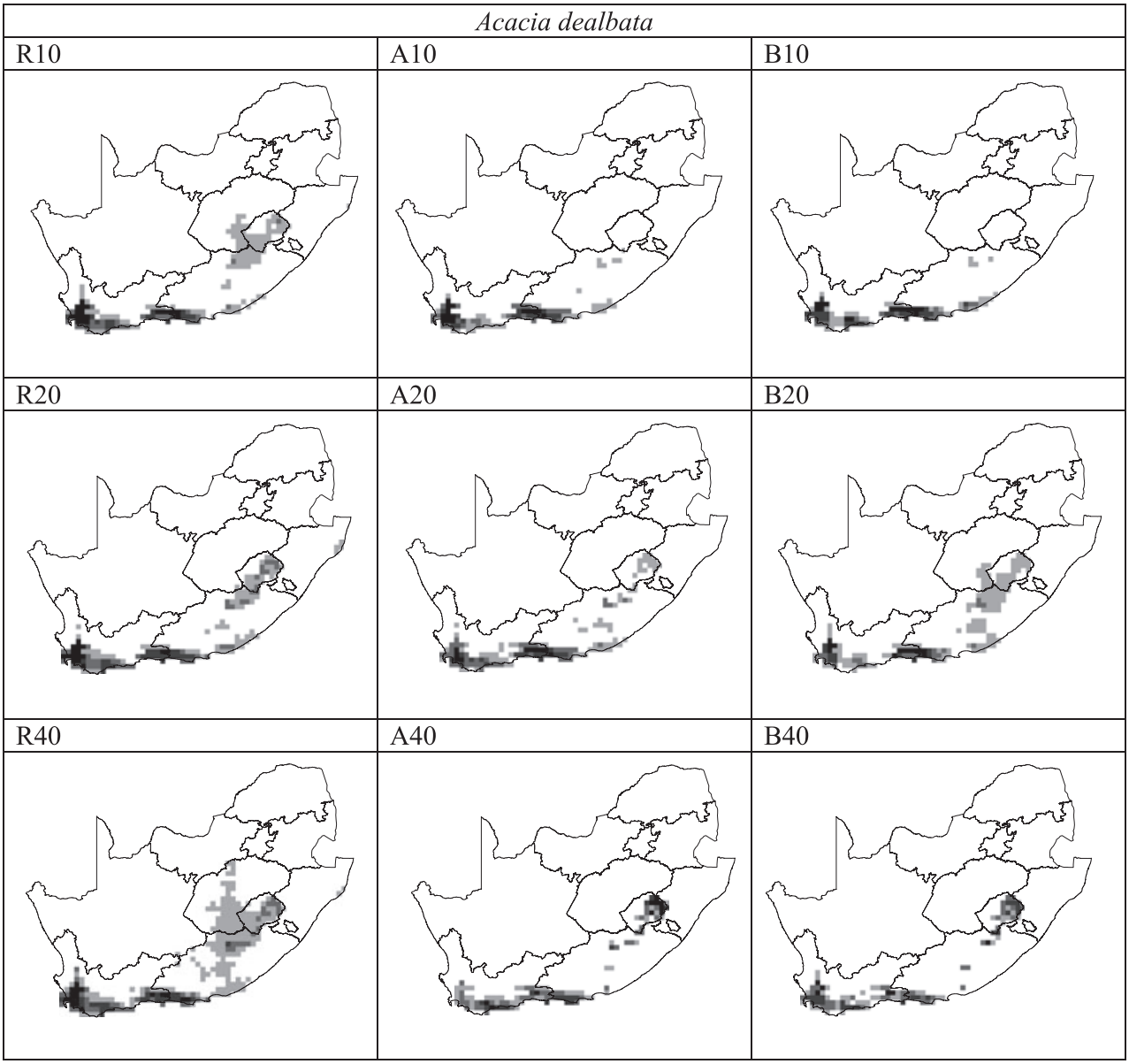

Figure 6 Introduced range predictions for *Acacia dealbata* with all three control datasets (R10, R20 and R40), datasets from treatment A (A10, A20 and A40) as well as datasets from treatment B (B10, B20 and B40).

environmental bias for any of the treatments analysed when we used correlation analysis. The sample size in this study ($n = 19$) may have been too small to detect a significant relationship given the amount of variation.

Environmental bias across treatments

There was a significant difference in environmental bias across treatments. However, there was also considerable environmental variation between treatments. As expected, bias was greater in the A40 and B40 treatments compared with the A10 and B10 treatments. This suggests that even though we did not find a significant relationship between geographical and environmental bias, the geographical bias simulated in the occurrence records resulted in some environmental bias. It is important to know whether the amount of environmental bias simulated here can be considered to be large or small. Steiner *et al.* (2008) compared the introduced niche position of two invasive ant species in terms of marginality values when modelled with different datasets. For one species the niche position was considered to be significantly different when the difference in the marginality value was 0.17. For the second species, niche position was considered to be significantly different when the difference in the marginality value was 0.3. The average difference in marginality values for the treatments analysed here was very small in comparison (largest difference approximately 0.007). Therefore, although the amount of environmental bias was significantly different between

treatments, the amount of bias present can be considered to be small.

Effect of bias on model performance

On average, for all species, geographical bias and underlying environmental bias did not significantly influence model performance. This is counterintuitive and in contrast to Kadmon *et al.* (2003), who found a significant decrease in model performance in the presence of environmental bias. An explanation for this result requires consideration of the ecological niches of the species concerned.

Several studies have reported niche shifts between the native and introduced ranges of alien invasive organisms (Mau-Crimmins *et al.*, 2006; Broennimann *et al.*, 2007; Fitzpatrick *et al.*, 2007; Beaumont *et al.*, 2009). This phenomenon has been recognized when an invasive species occupies habitats or climatic zones different from those encountered in the species' native range (Mau-Crimmins *et al.*, 2006; Broennimann *et al.*, 2007; Fitzpatrick *et al.*, 2007; Beaumont *et al.*, 2009). Broennimann *et al.* (2007) provide evidence of an invasive species occupying a climatically different niche in its native and introduced range. Similar results were found by Fitzpatrick *et al.* (2007), who showed that the invasive fire ant species initially invaded areas in their introduced range with conditions similar to that of their native range, and then started to spread to regions that were climatically dissimilar to their native range. These shifts in niche occupation were argued to be the result of either genetic change that occurred after introduction, for example hybridization (shifts in the fundamental niche), or release from biotic constraints, for example competitors (shifts in the realized niche), or a combination of these two factors (Mau-Crimmins *et al.*, 2006; Broennimann *et al.*, 2007; Fitzpatrick *et al.*, 2007; Beaumont *et al.*, 2009).

We calibrated models for 19 species using native range occurrence records and evaluated them with occurrence records obtained from their introduced range. To ensure that these evaluations were fair we had to assume that the niches occupied by these species were conserved between the native and introduced ranges (Wiens & Graham, 2005). Following Steiner *et al.* (2008), we tested for niche shift using ENFA (Hirzel *et al.*, 2002). Marginality values were not significantly different between the native and introduced ranges. Tolerance values showed significant differences for 12 out of the 19 species (Table 2). This indicated that for all the species considered here, no significant shift in the niche position had occurred. In other words, the centres of the niches occupied in the native and introduced ranges were in the same part of the environmental space, but the amount of environmental variation differed significantly between the native and introduced ranges.

To visualize the environmental space occupied by a species in its native and introduced range we plotted component scores of components 1 and 2 from a PCA performed on the environmental data associated with the occurrence records in the native and introduced range. Some species did show good overlap in the resulting environmental space, but the majority of species showed an expansion or contraction of the niche occupied in the introduced range (Appendix S2b).

The differences in environmental space that the species occupy in their native and introduced ranges could have limited our ability to detect significant effects of geographical bias on model performance. If the geographical bias that was simulated in the native range occurrence records resulted in bias in a different part of the environmental space from that occupied by the species in the introduced range then it would appear as if geographical bias was having no effect. Indeed, for most species the native range records excluded from the dataset (biased records) did not occur in the region of the environmental space occupied by the introduced range records that were used for model evaluation (Appendix S2b).

Our failure to detect a significant difference in model performance despite significant environmental bias across treatments may be explained by two factors: first, that the amount of environmental bias resulting from the simulated geographical bias was relatively small for most species; and second, that for many species the environmental bias simulated was not in that part of the environmental space occupied in the introduced range (where the model evaluation was undertaken).

For some species (e.g. *Acacia saligna* presented in Appendix S5), the predictions of potential distribution obtained when models were calibrated using biased datasets and projected to the introduced range were very similar to the predictions made using records from the corresponding control datasets. Distinct differences in the region predicted as suitable were, however, evident for a few species (e.g. *A. dealbata* presented in Fig. 6), suggesting that for some species the environmental gradients may not have been sufficiently sampled with the biased datasets. Loiselle *et al.* (2008) obtained similar results, and concluded that model performance can be quite good despite existing environmental bias, if models are calibrated with a sufficient number of occurrence records. This might explain our findings, since we used quite a large number of records for calibrating the models (Table 1). Perhaps the effect of bias would have been greater if fewer records were available for model calibration.

The evaluation of model performance is reliant on the quality of the testing set. We evaluated models with test sets that consisted of presence records obtained from the introduced range and pseudo-absence records that we generated. Pseudo-absences were randomly selected grid cells from the introduced map region that were not occupied by a presence record. A weakness of our approach is the implicit assumption that the records available are a true reflection of a species' distribution in the introduced range, and therefore that a species is at equilibrium with its environment and that sampling effort has been thorough. However, although we specifically selected species that were considered to be well established and well sampled in their introduced ranges, it is possible that some of the species may have been undersampled.

In conclusion, we found that geographical bias did not have a significant negative effect on the performance of ecological niche models for invasive plant species. Our results are novel in the sense that we focused specifically on invasive plant species where the environmental spaces occupied by the species in the native and introduced ranges may differ. When making predictions of introduced range for alien invasive species it may be necessary to calibrate models with native range records. This would most likely be the case when predicting the potential distribution of emerging invaders, when only a few introduced range records are available (Mgidi *et al.*, 2007), or when the risk of a potential invader needs to be assessed in areas where invasion has not yet occurred and therefore no introduced range records are available (De Meyer *et al.*, 2008). In this study we were able to demonstrate, at least for this suite of species, that model performance was not significantly reduced when geographical bias was present in the native range records used to calibrate the models. This could be because the amount of environmental bias resulting from the simulated geographical bias was relatively small for most species. However, for many species the environmental bias simulated was not in that part of the environmental space occupied in the introduced range (where the model evaluation was undertaken). This is probably why we did not find a significant reduction in model performance in the presence of environmental bias. Furthermore, the effects of sample size, spatial resolution of predictor variables, range size and geographical region may influence environmental bias and require further investigation. Loiselle *et al.* (2008) found that species with larger range sizes tended to be affected more by environmental bias than those with narrower ranges. Our study was conducted using relatively coarse-scale (15′) predictors, and finer-scale studies may yield different results. Differences in the environmental spaces occupied by a species between its native and introduced ranges may reduce the effect of geographical bias on model performance, especially when the bias is in a part of the environmental space that does not occur in the introduced range. Further studies on the relationship between geographical and environmental bias for invasive species would be valuable.

ACKNOWLEDGEMENTS

Lesley Henderson kindly provided the SAPIA database and assisted with the species selection process. Rod Randall provided information about the South African plants invading Australia. The Council of Heads of Australian Herbaria and the custodial institutions, the Australian National Herbarium (Canberra), the National Herbarium of New South Wales (New South Wales), Northern Territory Herbarium (Darwin and Alice Springs), Queensland Herbarium (Brisbane), State Herbarium of South Australia (Adelaide), Tasmanian Herbarium (Hobart), National Herbarium of Victoria (Melbourne) and the Western Australian Herbarium (Perth) are acknowledged for providing specimen data for Australian plants, and in particular, we thank Alex Chapman, Murray Fagg and Greg Whitbread for assistance. Paul Robins (Queensland Herbarium, Environmental Protection Agency), in charge of the HERBRECS, kindly provided distribution data. Many thanks to Steven Phillips and Miroslav Dudik for help and advice with MAXENT. We would also like to express our gratitude to Jeremy van der Wal, who provided statistical advice and shared some of his scripts with us. We thank Richard Pearson, Joaquín Hortal, Mathieu Rouget and two anonymous referees who provided helpful comments on an earlier draft of this manuscript. This project was funded by the University of Pretoria and the DST-NRF Centre of Excellence for Invasion Biology.

REFERENCES

Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43,** 1223–1232.

Beaumont, L.J., Gallagher, R.V., Thuiller, W., Downey, P.O., Leishman, M.R. & Hughes, L. (2009) Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions. *Diversity and Distributions*, **15,** 409–420.

Broennimann, O. & Guisan, A. (2008) Predicting current and future biological invasions: both native and invaded ranges matter. *Biology Letters*, **4,** 585–589.

Broennimann, O., Treier, U.A., Muller-Scharer, H., Thuiller, W., Peterson, A.T. & Guisan, A. (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters*, **10,** 701–709.

De Meyer, M., Robertson, M.P., Peterson, A.T. & Mansell, M.W. (2008) Ecological modelling and potential geographic distributions of Mediterranean fruit fly (*Ceratitis capitata*) and Natal fruit fly (*Ceratitis rosa*). *Journal of Biogeography*, **35,** 270–281.

Dennis, R.L.H. & Thomas, C.D. (2000) Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. *Journal of Insect Conservation*, **4,** 73–77.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29,** 129–151.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24,** 38–49.

Fitzpatrick, M.C., Weltzin, J.F., Sanders, N.J. & Dunn, R.R. (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography*, **16,** 24–33.

Funk, V.A. & Richardson, K.S. (2002) Systematic data in biodiversity studies: use it or lose it. *Systematic Biology*, **51,** 303–316.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8,** 993–1009.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135,** 147–186.

Henderson, L. (2007) Invasive, naturalized and casual alien plants in southern Africa: a summary based on the Southern African Plant Invaders Atlas (SAPIA). *Bothalia*, **37,** 215–248.

Hernandez, P.A., Graham, C.H., Masters, L.L. & Albert, D.L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29,** 773–785.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25,** 1965–1978.

Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002) Ecological-niche factor analysis: how to compute habitat suitability maps without absence data? *Ecology*, **83,** 2027–2036.

Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M. & Baselga, A. (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, **117,** 847–858.

Jiménez-Valverde, A. & Lobo, J.M. (2007) Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta Oecologia*, **31,** 361–369.

Kadmon, R., Farber, O. & Danin, A. (2003) A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications*, **13,** 853–867.

Kadmon, R., Farber, O. & Danin, A. (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, **14,** 401–413.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17,** 145–151.

Loiselle, B.A., Jorgenson, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G. & Montiel, O.M. (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography*, **35,** 105–116.

Mau-Crimmins, T., Schussman, H.R. & Geiger, H. (2006) Can the invaded range of a species be predicted sufficiently using only native-range data? Lehmann lovegrass (*Eragrostis lehmanniana*) in the southwestern United States. *Ecological Modelling*, **19,** 736–746.

McPherson, J.M. & Jetz, W. (2007) Effects of species' ecology on the accuracy of distribution models. *Ecography*, **30,** 135–151.

Mgidi, T., Le Maitre, D.C., Schonegevel, L., Nel, J.L., Rouget, M. & Richardson, D.M. (2007) Alien plant invasions – incorporating emerging invaders in regional prioritization: a pragmatic approach for Southern Africa. *Journal of Environmental Management*, **84,** 173–187.

O'Brien, E.M. (1993) Climate gradients in woody plant species richness: towards an explanation based on an analysis of southern Africa's woody flora. *Journal of Biogeography*, **20,** 181–198.

Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12,** 361–371.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, P. & Lees, D.C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33,** 1704–1711.

Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Peterson, A.T. (2007) Predicting species distribution from small occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34,** 102–117.

Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology*, **78,** 419–433.

Peterson, A.T., Papes, M. & Soberón, J. (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, **213,** 63–72.

Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31,** 161–175.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190,** 231–259.

Pimentel, D., McNair, S., Janecka, J., Wightman, J., Simmonds, C., O'Connell, C., Wong, E., Russel, L., Zern, J., Aquino, T. & Tsomondo, T. (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture, Ecosystems and Environment*, **84,** 1–20.

R Development Core Team (2004) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org.

Raes, N. & ter Steege, H. (2007) A null-model for significance testing of presence-only species distribution models. *Ecography*, **30,** 727–736.

Randall, R.P. (2002) *A global compendium of weeds*. Shannon Books, Melbourne.

Richardson, D.M. & Thuiller, W. (2007) Home away from home – objective mapping of high risk source areas for plant introductions. *Diversity and Distributions*, **1,** 1–14.

Robertson, M.P., Caithness, N. & Villet, M.H. (2001) A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. *Diversity and Distributions*, **7,** 15–27.

Robertson, M.P., Peter, C.I., Villet, M.H. & Ripley, B.S. (2003) Comparing models for predicting species potential distribution: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling*, **164,** 153–167.

Rouget, M., Richardson, D.M., Nel, J.L., Le Maitre, D.C., Egoh, B. & Mgidi, T. (2004) Mapping the potential ranges of major plant invaders in South Africa, Lesotho and Swaziland using climatic suitability. *Diversity and Distributions*, **10,** 475–484.

Roura-Pascual, N., Brotons, L., Peterson, A.T. & Thuiller, W. (2008) Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. *Biological Invasions*, **11,** 1017–1031.

Soberón, J. & Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2,** 1–10.

Steiner, F.M., Schlick-Steiner, B.C., Vanderwal, J., Reuther, K.D., Christian, E., Stauffer, C., Suarez, A.V., Williams, S.E. & Crozier, R.H. (2008) Combined modeling of distribution and niche in invasion biology: a case study of two invasive *Tetramorium* ant species. *Diversity and Distributions*, **14,** 538–545.

Vitousek, P.M., D'Antonio, C.M., Loope, L.L., Rejmánek, M. & Westbrooks, R. (1997) Introduced species: a significant component of human-caused global change. *New Zealand Journal of Ecology*, **21,** 1–16.

Welk, E. (2004) Constraints in range predictions of invasive plant species due to non-equilibrium distribution pattern: purple loosestrife (*Lythrum salicaria*) in North America. *Ecological Modelling*, **179,** 551–567.

Wiens, J.J. & Graham, C.H. (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution and Systematics*, **36,** 519–539.

Wilson, J.R.U., Richardson, D.M., Rouget, M., Procheş, Ş., Amis, M.A., Henderson, L. & Thuiller, W. (2007) Residence time and potential range: crucial considerations in modelling plant invasions. *Diversity and Distributions*, **13,** 11–22.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix S1 The 19 bioclimatic variables obtained from the WorldClim database that were used as predictor variables.

Appendix S2 The relationship between geographical and environmental bias in datasets for all the study species.

Appendix S3 The relationship between native and introduced range model performance.

Appendix S4 Environmental bias expressed as (a) difference in marginality and (b) difference in tolerance assessed across all treatments for all species.

Appendix S5 Introduced range predictions for *Acacia saligna*.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCHES

René Wolmarans has a Master's degree in zoology and is interested in ecological niche modelling and biological invasions.

Mark Robertson is an ecologist interested in the distributions of species. His research focuses on ecological niche modelling and predicting potential distributions of species of conservation concern and invasive species.

Berndt van Rensburg is interested in human-related impacts on biodiversity and macroecology, especially in southern Africa, and the conservation implications thereof.

Editor: Richard Pearson